

Not a Copy and Paste: Makings of Non-English Data Collection Instruments

FedCASIC 2019 Workshop

Lily Kapaku, Brianda Perez

Content, Language, and Data Products Branch

Decennial Census Management Division, U.S. Census Bureau, Washington, DC

Multilingual Automated Data Collection Instruments

Purpose

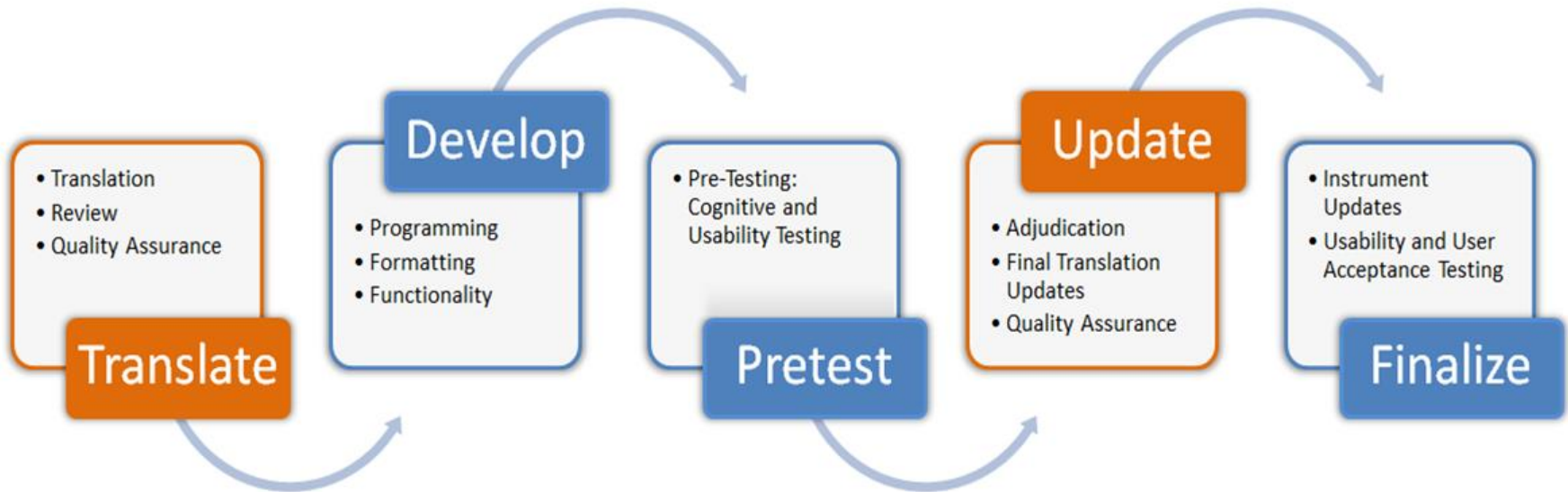
- The 2020 Census counts every person living in the United States, once, only once, and in the right place.
- Help reduce language barriers non-English speakers may encounter.
- Optimize the non-English content across data collection modes and operations.
- Provide translations that are accurate, meaningful and culturally relevant.

Multilingual Automated Data Collection Instruments

Background

Instrument/Material	2020 Non-English Support	2010 Non-English Support
<p>Internet Self-Response</p>	<p>12 Non-English Languages</p> <ul style="list-style-type: none"> Spanish, Chinese (Simplified), Vietnamese, Korean, Russian, Arabic, Tagalog, Polish, French, Haitian Creole, Portuguese, Japanese <p>Language Selection Criteria: 60,000 or more housing units</p>	<p>Not in Scope</p> <p><i>Fulfillment and Be Counted Questionnaires in 5 non-English languages, available via phone request or pickup: Spanish, Chinese (Simplified), Korean, Vietnamese, Russian</i></p> <p><i>Language Selection Criteria: 100,000 or more housing units</i></p>
<p>Census Questionnaire Assistance (Telephone) Interviewer Based</p>	<p>12 Non-English Languages</p> <ul style="list-style-type: none"> Spanish, Chinese (Mandarin and Cantonese), Vietnamese, Korean, Russian, Arabic, Tagalog, Polish, French, Haitian Creole, Portuguese, Japanese <p>Telecommunication Device for the Deaf also available</p> <p>Language Selection Criteria: 60,000 or more housing units</p>	<p>5 Non-English Languages</p> <ul style="list-style-type: none"> Spanish, Chinese (Mandarin and Cantonese), Korean, Vietnamese, Russian <p>Telecommunication Device for the Deaf also available</p> <p>Language Selection Criteria: 100,000 or more housing units</p>
<p>Enumerator Instrument (Mobile) Interviewer Based</p>	<p>Spanish (hand-held)</p>	<p>Not in scope</p>

Multilingual Automated Data Collection Instruments Process



Challenges and Findings

Style, Terminology, and Other Considerations

Challenges and Findings: Translations

Consider Multiple Meanings of English Terms

- English content with context specific translations

Translation of Vietnamese (race)

 Người Ấn độ
 Người Việt nam
 Người Hàn quốc

vs.

Vietnamese (language)



Challenges and Findings: Translations

English Terms That Should Be Included in Translations

- Some English terms are included because there is no equivalent term or because the English term is the optimal term.
- These English terms are emphasized using parentheses and/or italics

<input type="radio"/> Grandchild	<input type="radio"/> Nieto(a)
<input checked="" type="radio"/> Roommate or housemate	<input type="radio"/> <i>Roommate</i> o compañero(a) de casa
<input type="radio"/> Other	<input type="radio"/> Otro

- In a nursing home or group home
- In a jail or prison
- En un hogar de ancianos *o nursing*
- *home,* o en un hogar de grupo
- En una cárcel o prisión

Challenges and Findings: Translations

Consider Whether Responses Are Needed in English

- Some responses can be translated as necessary, others cannot
- Instruction to provide names in English added to address output requirements

What is your name and telephone number?

We will only contact you if needed for official Census Bureau business.

First Name	Middle Name	Last Name(s)
<input type="text"/>	<input type="text"/>	<input type="text"/>

귀하의 이름과 전화번호는 무엇입니까?
이름은 영문으로 입력해 주십시오. 인구조사국 업무 관계상 꼭 필요한 경우에만 연락을 드리겠습니다.

이름 (First Name)	중간 이름 (Middle Name)	성 (Last Name(s))
<input type="text"/>	<input type="text"/>	<input type="text"/>

Challenges and Findings: Translations

Translations Developed In Both Past And Future Tense

- Census Day is April 1, 2020. Respondents can start responding in March 2020.
- Verb tense changes dynamically based on when the respondent is completing the census.

Portuguese translation:

"will be living" = "**estará morando**"

"were living" = "**estava morando**"

Para coletar o seu endereço, primeiro precisamos saber onde você **estará morando** no dia 1 de abril de 2020.

Selecione onde **estará morando** no dia 1 de abril de 2020.

Para coletar o seu endereço, primeiro precisamos saber onde você **estava morando** no dia 1 de janeiro de 1970.

Você **estava morando** no endereço __?

Challenges and Findings: Translations

Mode-Specific Formality Considerations

- For interviewer instruments (telephone and mobile), mode-specific translations are crucial.
 - Need to incorporate politeness markers.

Spanish incorporates polite terms for you ("usted") and your ("su") instead of informal terms

A continuación, necesitamos saber la relación de cada persona con usted.

Muestre la pantalla o lea las opciones a la persona encuestada.

Mirando a la pantalla, Jane Doe es SU.

"How is SAM related to JIMMY? SAM is JIMMY's..."
Example shows 님 (Korean honorific) after Name fills.

SAM 님은 JIMMY 님과 어떤 관계 입니까?
SAM 님은 JIMMY 님의...

Challenges and Findings: Translations

Cannot Repurpose Translations

- Translations need to have 1-1 relationship with English text and be context specific.
- For example, cannot use a translation found in another question for “Other”
 - Другим родственником/родственницей is the translation of “Other” found on another screen, but is the equivalent of "Another relative/relative"

Including yourself, how many people will be living or staying at 101 MAIN ST on April 1, 2020?

1

2

Other

Сколько человек, включая Вас, будет проживать или находиться по адресу 101 MAIN ST 1 апреля 2020 г.?

1

2

Другим родственником/родственницей

Challenges and Findings: Translation and Design

Emphasizing Non-English Text

- There are instances where English text is emphasized in questionnaires in all capital letters.
 - Not all languages have capitals or emphasize text using capitals.
 - Substitute other appropriate styles, such as underlining, to ensure the emphasis is not lost.

Do NOT include:

- College students who live away from **<ADDRESS>** most of the year.
- Armed forces personnel who live away.

НЕ ВКЛЮЧАЙТЕ:

- Студентов колледжей, которые большую часть года не проживают по адресу **<ADDRESS>**.
- Военнослужащих, проживающих в другом месте.

لا تشمل:

- طلاب الجامعة الذين يسكنون بعيدا من **<ADDRESS>** معظم السنة.
- أفراد القوات المسلحة الذين يسكنون بعيدا.

Challenges and Findings: Translation and Design

Date of Birth

- When providing the date of birth, respondents may use different date formats (MM/DD/YYYY vs DD/MM/YYYY).
- Use numbers AND month names

01 - Janvyè
02 - Fevriye
03 - Mas
04 - Avril
05 - Me
06 - Jen
07 - Jiyè
08 - Out
09 - Septanm
10 - Oktòb
11 - Novanm
12 - Desanm

Haitian Creole

01 - styczeń
02 - luty
03 - marzec
04 - kwiecień
05 - maj
06 - czerwiec
07 - lipiec
08 - sierpień
09 - wrzesień
10 - październik
11 - listopad
12 - grudzień

Polish

01 - يناير / كانون الثاني
02 - فبراير / شباط
03 - مارس / آذار
04 - أبريل / نيسان
05 - مايو / أيار
06 - يونيو / حزيران
07 - يوليو / تموز
08 - أغسطس / آب
09 - سبتمبر / أيلول
10 - أكتوبر / تشرين الأول
11 - نوفمبر / تشرين الثاني
12 - ديسمبر / كانون الأول

Arabic

Challenges and Findings: Design

Response Option Design

- Use of X, can be interpreted as "No" or "Does not apply"
- Instead, use fill-in, checkmark or circle for affirmative response.

Radio button

What is JOHN DOE's sex?

Male

Female

ما هو جنس JOHN DOE؟

ذكر

أنثى

Check box

Is JOHN DOE of Hispanic, Latino, or Spanish origin?

No, not of Hispanic, Latino, or Spanish origin

Yes, Mexican, Mexican American, Chicano

Yes, Puerto Rican

Yes, Cuban

Yes, another Hispanic, Latino, or Spanish origin

Enter, for example, Salvadoran, Dominican, Colombian, Guatemalan

هل JOHN DOE من أصل هسباني أو لاتيني أو إسباني؟

لا ليس من أصل هسباني أو لاتيني أو إسباني

نعم، مكسيكي، مكسيكي أمريكي، تشيكانو

نعم، بورتوريكي

نعم، كوبي

نعم، من أصل آخر هسباني أو لاتيني أو إسباني

ادخل، على سبيل المثال، سلطادوري، دومينيكي، كولومبي، غواتيمالي، إسباني، إكوادوري، إلخ.

Challenges and Findings: Design

Optimize How Translations Appear

- Text expansion causes translations to appear differently than the English content.
- Translations need additional work to be optimized for non-English data collection instruments.

Name translations
without html code

Имя (First Name)	Второе имя или отчество (Middle Name)	Фамилия(и) (Last Name (s))
------------------	------------------------------------------	-------------------------------

Russian internet instrument
with name field translations
(with and without
 code)

Имя (First Name)	Второе имя или отчество (Middle Name)	Фамилия(и) (Last Name(s))
-----------------------	--------------------------------------------	--------------------------------

Имя (First Name)	Второе имя или отчество (Middle Name)	Фамилия(и) (Last Name(s))
<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="button" value="Удалить"/>		

Challenges and Findings: Design

Optimizing Non-English Buttons

- Length of translation should also be assessed for optimal design
 - Example: vary font and size in buttons
- Directional elements should incorporate correct directions
 - Arrows for languages read right-to-left, such as Arabic: right for “Previous” and left for “Next”

Japanese



Russian



Chinese



Arabic

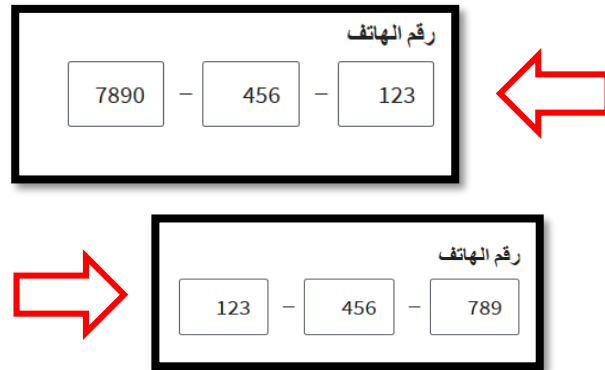


Challenges and Findings: Design

Optimize Field Order and Layout

- Though Arabic is read right to left, phone numbers are usually read left to right.
- Need to correctly place 4-digit field.

Respondents entering (123) 456-7890
might insert (789) 456-123X



Best Practices Summary

- **Translations:**
 - Use professional translators and follow industry best practices for quality assurance.
 - Use plain language and provide translator notes to ensure English meaning is clear.
 - Include English terms in the translation when needed for clarity.
 - Consider how the translation sounds in spoken modes, and incorporate appropriate formal terms.
 - Consider dynamic content needs (fills, singular/plural, verb tense adjustments)
 - Do not "repurpose" translations throughout an instrument. Each translation must have a context-specific relationship with the English text.
 - Use usability and cognitive testing to identify critical issues.
- **Design**
 - Consider how the translation looks (directional text, size, line breaks, formatting)
 - Consider whether visual elements might be misunderstood (example: marking responses with an X).

Fragen? Ερωτήσεις? ʻāqama? प्रश्नहरू? Àwọn ibèèrè?
 ?תלמש Вопросы? கேள்விகள்? Հարցերի प्रश्न? Kérdések
 ప్రశ్నలు? Pytania 问题? Pertanyaan?
 問題? Sorular? Questions? Câu hỏi? ਪ੍ਰਸ਼ਨ?
 Domande? Kesyon? ¿Preguntas? 質問? Întrebări?
 سؤال? Otázky? Mga tanong? Perguntas? 질문? Питання?
 ಹಿರಿಶಿಶು? Klausimai الأسئلة? Вопросы ചോദ്യങ്ങൾ? Pitanja?
 Pyetje? Vragen? Ajuju? അപ്രകാശം? سوالات? សំណួរ? កំណាម?

Contact Information

Lily Kapaku
Content, Language, and Data Products Branch
Decennial Census Management Division
U.S. Census Bureau
301-763-0189
lily.kapaku@census.gov

Brianda Perez
Content, Language, and Data Products Branch
Decennial Census Management Division
U.S. Census Bureau
301-763-5536
Brianda.Perez@census.gov

2020 Census Connect with Us



Sign up for and manage alerts at
<https://public.govdelivery.com/accounts/USCENSUS/subscriber/new>



More information on the 2020
Census:
<http://www.census.gov/2020Census>



[facebook.com/uscensusbureau](https://www.facebook.com/uscensusbureau)



twitter.com/uscensusbureau



[youtube.com/user/uscensusbureau](https://www.youtube.com/user/uscensusbureau)



[instagram.com/uscensusbureau](https://www.instagram.com/uscensusbureau)



[pinterest.com/uscensusbureau](https://www.pinterest.com/uscensusbureau)